Skill Demand and Wages: Evidence from Online Job Posts in Austria

Lennart Ziegler

University of Vienna

Arbeitsmarktworkshop - 9 November 2018

Motivation

Technological progress drives substantial changes in the demand for skills

- What skills are currently needed on the labor market?
- Leading theory of skill-biased technological change (SBTC):
  - Computerization $\Rightarrow$ Non-routine analytic and interactive tasks $\uparrow$ $\Rightarrow$ Analytic and managerial skills $\uparrow$

<u>Common approach to measure skill demand</u>: Infer profile from occupational dictionaries

- Comprehensive characterization but only periodically revised & not available for many countries

Contribution

This paper: Infer skill requirements from online job posts

Analysis of skill demand and returns in Austrian labor market

1. Identify most frequent skill requirements
2. Estimate returns to different types of skills
3. Quantify spatial differences in skill requirements and contribution to urban-rural wage gap

Methodological contribution: Online job posts as source in labor market research

- How much information about skills do job ads contain?
- Are these skill measures able to explain wage differences?
- Focus on interpretation, limitations, measurement error

Literature

Returns to cognitive and social skills

- Test scores (Bowles, 2001; Heckman et al., 2006; Lindqvist & Vestman, 2010)
- Survey data (Hanushek et al., 2015)
- Occupational dictionaries (Autor & Handel, 2013)
- Job ads (Kahn & Deming, 2018)

Spatial differences in skill demand/returns

- Urban-rural wage gap in many countries (Moretti, 2010)
- Determinants: Worker migration (Glaeser & Mare, 2001), human capital spillovers (Moretti, 2004), amenities (Chen & Rosenthal, 2008)

Wage setting in Austria

The vast majority of Austrian workers (>95%) is covered by collective bargaining agreements (CBAs)

- Most are industry-specific, some firm-level agreements
- No collective bargaining in a few sectors

Working time, wages and other benefits (holiday pay etc.) are specified for each profession

- Pay scale often based on age and experience
- Collective bargaining wage defines legally binding lower bound
- Employers may overpay depending on qualification and experience

Wage posting in job ads

As of March 2011, employers are required to report the CBA wage in job postings

- Since August 2013, also in sectors without collective bargaining agreement
- Employers may be fined for violation by local authorities
- Posted wage should exclude bonus or other extra payments
- Unit of time must be reported (hour, month or year)

Examples:

- *"[...] offers a payment of 2.757,40 EUR (gross) per month (14 times p.a.) [...]"*
- "*We offer you a gross annual wage starting from EUR 40,000 and depending on your qualification and experience.*"

Posted wages and skill requirements

Final wage depends on qualification of applicant:

$$\text{Actual wage} = \underbrace{\text{Posted wage}}_{f(\text{Min. qualification})} + \underbrace{\text{Overpay}}_{\text{g(Actual qualification - Min. qualification)}}$$

- Match with skill requirements $\uparrow \Rightarrow$ Qualification $\uparrow$
- $\frac{\partial \text{Actual wage}}{\partial \text{Skill}} > \frac{\partial \text{Posted wage}}{\partial \text{Skill}}$

$\Rightarrow$ Skill returns to posted wages are lower bound for actual returns

Posted skills and skill content

Job ads can only give rough profile description

- Compared to occupational dictionaries, focus on different scope of skills
  - More emphasis on soft skills (e.g flexibility)
  - Less emphasis on hard skills (e.g. numeracy skills)
- Difficult to describe relative importance of skills
- Emphasis on major skills, omittance of minor skills

Posted skills and skill content

Rationale to omit minor skills:

1. Not discourage otherwise suitable job applicants
2. Put implicitly more weight on major skills

Example: Programmers and social skills

Let s_{ki} be a continuous relevance measure of skill k

- Model with occupational dictionaries: $wage_i = \sum_{k=1}^{K} \alpha_k s_{ki}$
- Model with job ad descriptions: $wage_i = \sum_{k=1}^{K} \beta_k I(s_{ki} \geq \underline{s_k})$

Data

Job posts scraped from *karriere.at* every 2-4 weeks (June - October 2018)

- Portal claims to be “biggest career website in Austria”

Skills and wages inferred from individual ad text:

- Skills matched on keywords
- Wages converted to monthly earnings

Additional information from ad classification:

- Experience required (yes/no)
- Type of work (regular, vocational training, internship)
- Extent of work (full-time/part-time)
- Days since ad posted
- Industry group

Data

Sample restrictions:

- Full-time positions
- Internships and vocational training excluded
- Job ads without posted wages or outside Austria

$\rightarrow$ Sample size: 41,374 ads

Job ad example

universität wien

AML Data Analyst (m/f)

Paysafe

Wien · Vollzeit · Berufserfahrung · 16.10.2018

Firmenprofil ansehen

Paysafe is a leading global provider of end-to-end payment solutions. Our core purpose is to enable businesses and consumers to connect and transact seamlessly through our industry-leading capabilities in payment processing, digital wallets and online cash solutions.

Your Key Responsibilities

- Analyse trends and transaction movements from a quantitative perspective
- Close cooperation with other departments (e.g. marketing, payment network, VIP agents, ...) to gain an explanation for certain trends and movements
- Investigate escalated cases where complex ML suspicions have been identified as well as customer

Job ad example

Your profile

- A-levels or equivalent, with at least 2 years' professional experience preferably in Fraud & Anti Money Laundering, Risk Assessment or data analysis
- Ability to investigate at database level and usage of BI tools (e.g. Tableau) as well as basic skills in SQL
- Deep knowledge of meta data and technical parameters at different kinds of network communication (http, https, SSL, TCP, UDP, RDP, VPN…) or anonymization tools like TOR
- Knowledge of ongoing threads regarding cybercrime, e.g. malware, ransomware, botnets, different exploit rootkits and other IT system vulnerabilities
- Powers of deduction to translate manual processes to automated processes
- Experience with prepaid products, e-wallet payment solutions or crypto currencies desirable
- Excellent MS-Office knowledge (especially Excel)
- German skills and fluent English skills required
- Strong analytical skills with an eye for detail as well as affinity for numbers
- Independent, collaborative and pro-active working style
- Results-focused with good communication skills

We Offer

- Challenging and rewarding environment to develop your skills and expertise further
- A fast paced and evolving business with international background
- Personal atmosphere where successfully reached goals get celebrated at team events
- A modern office providing an open and friendly working atmosphere
- International development and career opportunities
- Flexi-time, mobile working and other great company benefits
- An annual gross salary starting from € 30.000,00 with a view to increase based on the qualification and experience

If you are interested in joining our team,
please apply with your CV and cover letter online.

Take a look and visit us!

Sample statistics

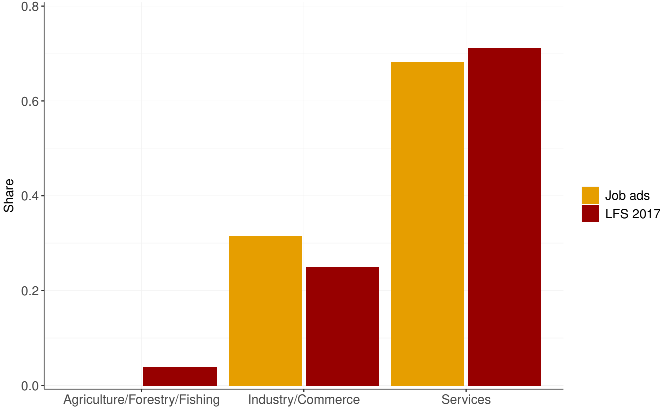
	Mean	Std. Dev.	Observations
Job post characteristics			
# words per ad	266.57	95.55	41,374
Days online	5.29	3.79	41,374
For beginners	0.04	0.19	41,374
College	0.24	0.43	41,374
Monthly salary	2,701.85	866.08	41,374
Firm Characteristics			
1-10 employees	0.11	0.31	34,077
11-100 employees	0.27	0.44	34,077
101-500 employees.	0.28	0.45	34,077
>500 employees	0.34	0.47	34,077
Firm age (in years)	47.25	62.29	14,074

Table: Job posts and firm characteristics

Representativeness

Job ads *versus* universe of vacancies

- Online job posts tend to over-represent white collar jobs
- Some vacancies are filled internally



Text-pattern analysis

Methodology:

- Identify skill requirements based on text pattern in job description

Procedure:

1. Rank words in job description by overall frequency
2. Extract keywords that describe job skills
3. Group words into skill categories

Text-pattern analysis

Skill group	Specific skill	Examples
Hard skills	Analytical skills	Problemlösungskompetenz, analytisches Denken
	Programming	Programmierer, Programming, Python, SQL
	MS-Office skills	MS Office, MS-Office, Microsoft Office
	Foreign language	Englisch, Französisch, Italenisch
Communication skills	Communication skills	Kommunikationsfähigkeit, kommunikativ
Managerial skills	Entrepreneurial skills	Unternehmerisches Denken/Geist
	Leadership skills	Führungsstärke, Führung
Other soft skills	Teamwork	Teamwork, gerne im Team arbeiten, Teamplayer
	Organizational skills	Organisationstalent, Organisationsfähigkeit
	Self-reliance	Eigeninitiative, eigenverantwortlich
	Assertiveness	Durchsetzungsvermögen
	Creativity	Kreativität, kreativ
	Stress tolerance	Belastbarkeit, Stressresistenz, Stress
	Reliability	Zuverlässigkeit, Verlässlichkeit

Number of skills

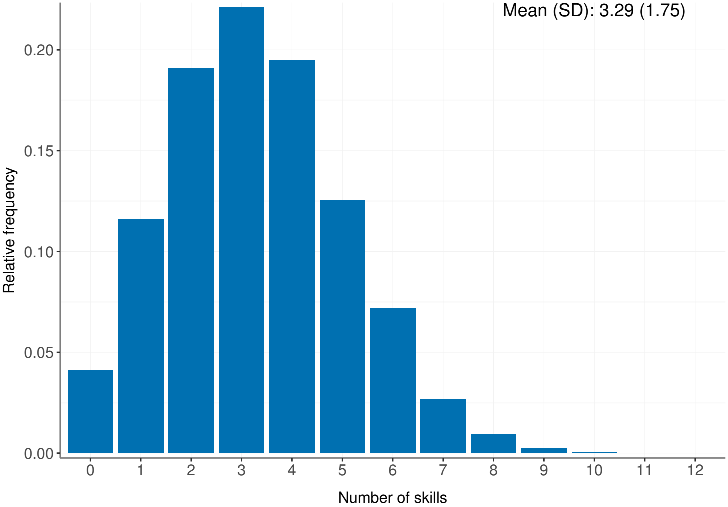


Figure: Frequency of skills per job post

Skill requirements

Skill types			
	Share		Share
Communicative	0.48	Leadership	0.18
Language	0.48	Stress-tolerant	0.15
Teamwork	0.40	Programming	0.14
Self-reliant	0.34	Organized	0.10
Analytical	0.32	Creative	0.09
MS-Office skills	0.28	Assertive	0.07
Reliable	0.20	Entrepreneurial	0.05
Observations: 41,374			

Table: Relative frequency of skill types

Wage regressions

$$log(wage_{ij}) = S_{ij}^{'}\beta + X_{ij}^{'}\gamma + \delta_j + u_{ij}$$

- S_{ij}: Skill measure of ad i of firm j
  1. Number of skills ($\#skills_i \in [0, 14]$)
  2. Vector of skill type indicators
- X_{ij}: Control variables (Beginner indicator, days online, college degree required)
- δ_j: Firm fixed effects

Posted wages determined by productivity and industry/firm bargaining power

$\Rightarrow$ Firm fixed effects account for differential bargaining power

Number of skills

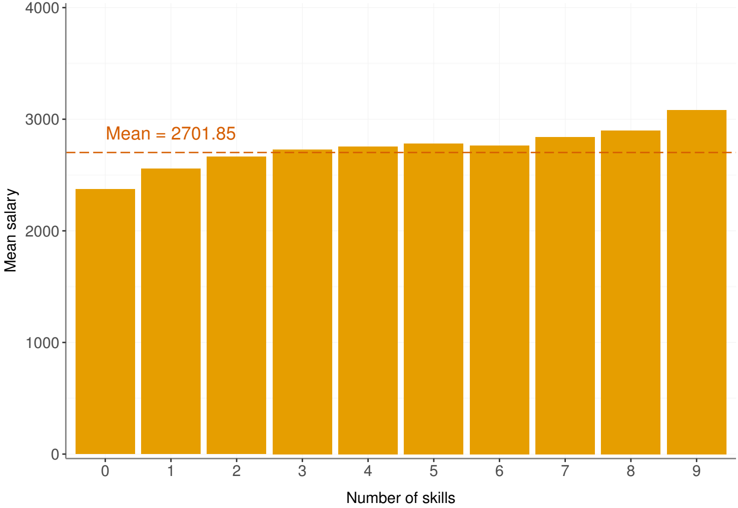


Figure: Wage by number of job-ad skills (Note: Last group contains ads with ≥ 9 skills)

Number of skills

universität wien

	(1)	(2)	(3)	(4)
# skills	0.020***	0.012***	0.017***	0.013***
	(0.001)	(0.001)	(0.001)	(0.001)
College		0.160***		0.119***
		(0.003)		(0.003)
Beginner		-0.174***		-0.158***
		(0.007)		(0.007)
Days online		0.001**		0.000
		(0.000)		(0.000)
Constant	7.793***	7.779***	7.586***	7.587***
	(0.003)	(0.004)	(0.153)	(0.148)
Firm FE			✓	✓
Adj. R^2	0.014	0.078	0.455	0.486
Observations	41,374			

Note – Standard errors are reported in parentheses. * significant at 10% level, ** significant at 5% level, *** significant at 1% level.

Table: Impact on logarithm of wages

Skill types

	(1)		(2)	
Hard skills				
Analytical	**0.098**	(0.003)	**0.053**	(0.003)
Programming	**0.087**	(0.004)	**0.033**	(0.003)
MS-Office skills	**-0.079**	(0.003)	**-0.065**	(0.003)
Language	**0.083**	(0.003)	**0.043**	(0.003)
Managerial	**0.146**	(0.003)	**0.131**	(0.003)
Soft skills				
Communicative	**0.024**	(0.003)	**0.017**	(0.002)
Other soft skills	**-0.050**	(0.003)	**-0.022**	(0.003)
Firm FE & controls			✓	
Adj. R-Squared	0.138		0.530	
Observations	41,374			

Note – Standard errors are in parentheses. **Bold** coefficients are significant at the 1%-level. Control variables include *college*, *beginner* and *days online*.

Table: Impact on logarithm of wages

Measurement error

Text-pattern analysis may suffer from measurement error

- **Under-detection**: Paraphrasing skill instead of naming
- **Over-detection**: Wrongfully attribute characteristic to applicant profile

What is the potential bias due to measurement error?

- Assumption: Measurement error is classical
- Notation: Measured skill $\tilde{S}_i$ and actual skill indicator S_i
- Under-detection probability: $r_u = P(\tilde{S}_i = 0 | S_i = 1)$
- Over-detection probability: $r_o = P(\tilde{S}_i = 1 | S_i = 0)$

True effect:

$$\beta = E(y_i|S_i = 1) - E(y_i|S_i = 0)$$

Measured effect:

$$\begin{aligned} b &= E(y_i|\tilde{S}_i = 1) - E(y_i|\tilde{S}_i = 0) \\ &= [1 - P(S_i = 0|\tilde{S}_i = 1) - P(S_i = 1|\tilde{S}_i = 0)] \times \beta \\ &= (1 - \frac{r_o \times [\frac{1-r_u-P(\tilde{S}_i=1)}{1-r_u-r_o}]}{P(\tilde{S}_i = 1)} - \frac{r_u \times [\frac{P(\tilde{S}_i=1)-r_o}{1-r_u-r_o}]}{1 - P(\tilde{S}_i = 1)}) \times \beta \end{aligned}$$

Measurement error

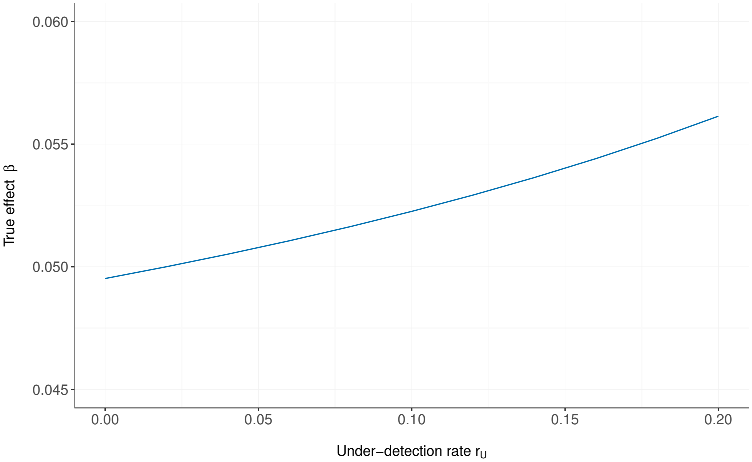


Figure: Potential measurement error (Example: Analytical skills ($r_o = 0$))

Spatial analysis

Analysis of spatial differences in skill demand and returns

1. Compare skill requirements between rural and urban districts
2. Decompose urban-rural wage gap into level and return differences in skills

$\Rightarrow$ Ads classified as **urban** if workplace in districts *Vienna*, *Linz*, *Graz*, *Salzburg*, *Innsbruck*, *Klagenfurt* ($\rightarrow$ 63% of job ads)

Job ads across Austria

universität wien

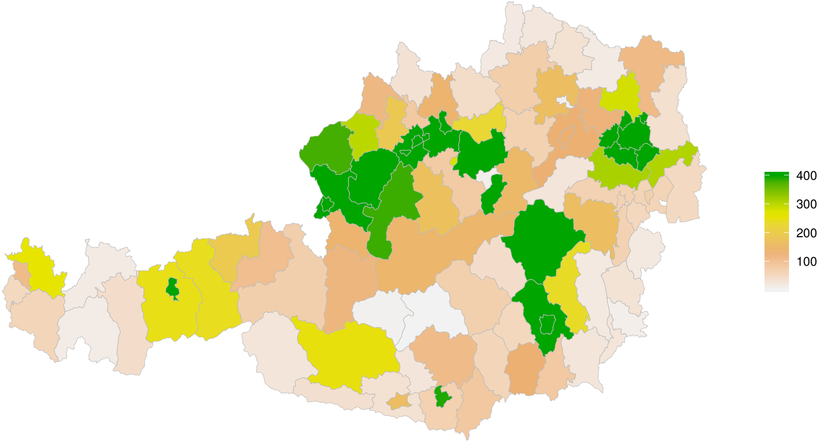


Figure: Frequency of job ads by district (Note: Highest frequency group refers to 400 ads or more.)

Urban-rural skill shares

Urban		Rural	
Skill	**Share**	**Skill**	**Share**
Communicative	0.51	Communicative	0.44
Language	0.51	Language	0.44
Teamwork	0.39	Teamwork	0.41
Self-reliant	0.35	Self-reliant	0.32
Analytical	0.35	MS-Office	0.28
MS-Office	0.28	Analytical	0.27
Leadership	0.19	Reliable	0.23
Reliable	0.18	Leadership	0.18
Programming	0.16	Stress-tolerant	0.15
Stress-tolerant	0.14	Programming	0.11
Organized	0.10	Organized	0.10
Creative	0.10	Creative	0.08
Assertive	0.06	Assertive	0.07
Entrepreneurial	0.06	Entrepreneurial	0.05
Mean skills: 3.37		Mean skills: 3.14	
N=25,862		N=15,512	

Wage decomposition

Separate estimation of wage-regression by regions:

$$log(wage_{gi}) = \beta_{g0} + \sum_{k=1}^{14} \beta_{gk} S_{gki} + u_{gi},\ g = U, R$$

Rewrite total wage gap $\hat{\Delta} = \overline{log(wage_U)} - \overline{log(wage_R)}$ as

$$\hat{\Delta} = \underbrace{(\hat{\beta}_{U0} - \hat{\beta}_{R0}) + \sum_{k=1}^{14} (\hat{\beta}_{Uk} - \hat{\beta}_{Rk})\bar{S}_{Uk}}_{\text{Wage structure effect}} + \underbrace{\sum_{k=1}^{14} (\bar{S}_{Uk} - \bar{S}_{Rk})\hat{\beta}_{Rk}}_{\text{Composition effect}}$$

where *Rural* serves as base group.

Alternatives: (i) Urban as base group, (ii) Weighted by relative group size

Wage decomposition

	Total	Constant	Skill returns	Skill levels
Base: Rural				
Estimate	0.082	0.092	-0.029	0.019
Standard error	(0.003)	(0.006)	(0.005)	(0.001)
Base: Urban				
Estimate	0.082	0.092	-0.032	0.023
Standard error	(0.003)	(0.006)	(0.005)	(0.001)
Base: Weighted				
Estimate	0.082	0.092	-0.031	0.021
Standard error	(0.003)	(0.006)	(0.005)	(0.001)

Note – Bootstrapped standard errors are in parentheses. In the weighted decomposition, return differences are $\sum_{k=1}^{14}(\tilde{\beta}_k - \hat{\beta}_{0k})\bar{S}_{0k} + \sum_{k=1}^{14}(\hat{\beta}_{1k} - \tilde{\beta}_k)\bar{S}_{1k}$ and level differences are $\sum_{k=1}^{14}(\bar{S}_{1k} - \bar{S}_{0k})\tilde{\beta}_{0k}$, where $\tilde{\beta}_k = p\hat{\beta}_{0k} + (1-p)\hat{\beta}_{1k}$ and p is the relative size of the base group.

Conclusion

Online job ads allow to quantify skill demand and returns

- Skill requirements have substantial explanatory power in wage regressions (net of firm FE)
- Hard skills are associated with higher returns than soft skills
- Measurement error unlikely to explain differences

Demand differences in key skill requirements between urban and rural districts

- Wage gap decomposition suggests higher skill return but lower *valuable* skill levels in rural areas

Skill correlation matrix

	Comm.	Ana.	Creat.	Team.	Self-rel.	Asser.	Org.	Str.-tol.	Rel.	Entr.	Lead.	Progr.	Lang.
Ana.	**0.12**												
Creat.	**0.05**	**0.05**											
Team.	**0.07**	**0.02**	0.01										
Self-rel.	**0.07**	**0.04**	**0.06**	**0.05**									
Asser.	**0.10**	**0.07**	0.00	**0.02**	**0.04**								
Org.	**0.11**	**-0.03**	**0.03**	**0.04**	**0.04**	**0.06**							
Str.-tol.	**0.02**	**-0.04**	**-0.02**	**0.10**	**0.03**	**0.07**	**0.10**						
Rel.	**-0.04**	**-0.07**	**-0.02**	**0.08**	0.01	0.00	**0.05**	**0.09**					
Entr.	**0.07**	**0.06**	0.01	**-0.03**	**0.02**	**0.08**	**0.03**	0.00	**-0.03**				
Lead.	**0.06**	0.00	-0.00	**-0.05**	**0.02**	**0.11**	**0.09**	0.01	**-0.02**	**0.12**			
Progr.	**-0.02**	**0.14**	**0.05**	**0.02**	**-0.03**	**-0.07**	**-0.09**	**-0.08**	**-0.07**	**-0.06**	**-0.11**		
Lang.	**0.14**	**0.17**	**0.03**	0.01	-0.00	**0.04**	**0.04**	**-0.02**	**-0.09**	0.01	0.01	**0.08**	
Admin.	**0.07**	**-0.02**	-0.01	**0.06**	**0.08**	**0.07**	**0.13**	**0.10**	**0.07**	**0.02**	**0.03**	**-0.18**	**0.06**

Table: Correlation matrix (**bold**: significant at 1%-level)

Specific skill types (1)

	(1)		(2)	
Analytical	**0.091**	(0.003)	**0.050**	(0.002)
Programming	**0.086**	(0.004)	**0.034**	(0.003)
MS-Office	**-0.073**	(0.003)	**-0.062**	(0.003)
Language	**0.081**	(0.003)	**0.042**	(0.003)
Entrepreneurial	**0.097**	(0.006)	**0.092**	(0.005)
Leadership	**0.142**	(0.003)	**0.129**	(0.003)
Organized	**-0.044**	(0.004)	**-0.028**	(0.004)
	[...]		[...]	
Firm FE & controls			✓	
Adj. R-Squared	0.160		0.541	
Observations	41,374			

Note – Standard errors are in parentheses. **Bold** coefficients are significant at the 1%-level.

Specific skill types (2)

	(1)		(2)	
	[...]		[...]	
Communicative	**0.021**	(0.003)	**0.014**	(0.003)
Teamwork	**-0.029**	(0.003)	**-0.015**	(0.002)
Creative	-0.001	(0.005)	0.011	(0.005)
Self-reliant	**0.011**	(0.003)	0.005	(0.002)
Assertive	**0.076**	(0.005)	**0.064**	(0.004)
Stress-tolerant	**-0.072**	(0.004)	**-0.030**	(0.003)
Reliable	**-0.046**	(0.003)	**-0.037**	(0.003)
Firm FE & controls			✓	
Adj. R-Squared	0.160		0.541	
Observations	41,374			

Note – Standard errors are reported in parentheses. **Bold** coefficients are significant at the 1%-level.

Skill interactions

	(1)	(2)	(3)	(4)
Analytical	0.091***	0.097***	0.050***	0.053***
	(0.003)	(0.004)	(0.002)	(0.004)
Communicative	0.021***	0.024***	0.014***	0.016***
	(0.003)	(0.003)	(0.002)	(0.003)
Analytical × Communicative		-0.011*		-0.006
		(0.006)		(0.005)
Intercept	7.783***	7.781***	7.612***	7.612***
	(0.003)	(0.003)	(0.140)	(0.140)
Firm FE & controls			✓	✓
Adj. R-Squared	0.160	0.160	0.541	0.541
Observations	41,374			

Note – Standard errors are reported in parentheses. * significant at 10% level, ** significant at 5% level, *** significant at 1% level. All regressions control for remaining skill types. Other control variables are *college*, *beginner* and *days online*.

Wage decomposition

	Base: Rural		Base: Urban		Weighted	
	Coeff.	SE	Coeff.	SE	Coeff.	SE
Total	0.082	(0.003)	0.082	(0.002)	0.082	(0.003)
Constant	0.085	(0.005)	0.085	(0.006)	0.085	(0.006)
Return differences						
Skills	-0.022	(0.005)	-0.026	(0.005)	-0.024	(0.005)
College	-0.010	(0.002)	-0.007	(0.001)	-0.008	(0.001)
Level differences						
Skills	0.016	(0.001)	0.020	(0.001)	0.019	(0.001)
College	0.013	(0.001)	0.009	(0.001)	0.011	(0.001)

Note – Bootstrapped standard errors are in parentheses. In the weighted decomposition, return differences are $\sum_{k=1}^{K}(\tilde{\beta}_k - \hat{\beta}_{0k})\bar{X}_{0k} + \sum_{k=1}^{K}(\hat{\beta}_{1k} - \tilde{\beta}_k)\bar{X}_{1k}$ and level differences are $\sum_{k=1}^{K}(\bar{X}_{1k} - \bar{X}_{0k})\tilde{\beta}_{0k}$, where $\tilde{\beta}_k = p\hat{\beta}_{0k} + (1-p)\hat{\beta}_{1k}$ and p is the relative size of the base group.